DOCUMENT RESUME

ED 411 862 IR 056 689

AUTHOR	Karen, Vladimir; Psohlavec, Stanislav
TITLE	Preservation and Access to Manuscript Collections of the Czech National Library.
PUB DATE	1996-00-00
NOTE	6p.; In: Online Information 96. Proceedings of the International Online Information Meeting (20th, Olympia 2, London, England, United Kingdom, December 3-5, 1996); see IR 056 631.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Access to Information; *Archives; Computer Software; *Electronic Libraries; Foreign Countries; Internet; *Library Collections; Library Materials; *National Libraries; Optical Data Disks; *Preservation; Printed Materials
IDENTIFIERS	*Czech Republic; *Digitizing; Historical Materials; HTML; UNESCO

ABSTRACT

In 1996, the Czech National Library started a large-scale digitization of its extensive and invaluable collection of historical manuscripts and printed books. Each page of the selected documents is scanned using a high-resolution, full-color digital camera, processed, and archived on a CD-ROM disk. HTML coded description is added to the entire document and to each page according to proposed recommendations currently under evaluation by the UNESCO Memory of the World Program Subcommittee for Technology. Comfortable and easy-to-use Windows software, ManuFret, is available for searching and viewing the manuscripts at a higher level. The collection of digitized documents is now open to the public in the National Library's manuscript study room and copies of the CD-ROMs are offered to other libraries world-wide. The catalog of digitized documents is available on the Internet. Internet accessibility to the full-text of all documents is being investigated. The paper addresses preservation and access issues, and describes the current and future status of the project. It also briefly discusses implications for other libraries that are thinking of digitizing their collection. (Author/SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 411 862

Preservation and access to manuscript collections of the Czech National Library

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
B.P. Jeapes
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Vladimír Karen and Stanislav Psohlavec
Albertina icome Praha s.r.o., Czech Republic

Abstract: *The Czech National Library started a large-scale digitisation of its extensive and invaluable collection of historical manuscripts and printed books in 1996. Each page of the document selected is scanned using a high-resolution digital camera in full-colour, processed and archived on a CD-R discs. HTML coded description is added to the entire document and to each page according to proposed recommendations currently under evaluation by the UNESCO Memory of the World Programme Sub-committee for Technology. Comfortable and easy-to-use Windows software, ManuFret, is available for searching and viewing the manuscripts at a higher level.*
The manuscript digitisation programme dates back to 1993 when the first pilot project was started and the 'Memory of the World CD-ROM' was published. Since then, several CD-ROMs have been published. Albertina icome Praha s.r.o. has been a partner of the Czech National Library since the very beginning of the programme and has actively participated in the preparation of the UNESCO recommendations.
The collection of digitised documents is now open to the public in the National Library's manuscript study room and copies of the CD-ROMs are offered to other libraries world-wide on demand. The catalogue of the digitised documents is available on the Internet and Internet accessibility to full documents is being investigated. The referee will address technical aspects of manuscript digitisation and publishing that have been solved through the years, and software solutions will be demonstrated.

Keywords: digitisation, old manuscripts, history, CD-ROM, Internet, UNESCO

1. Memory of the World UNESCO programme

The Memory of the World Programme of the UNESCO General Information Programme dates back to 1992, when the first aims of the Programme were defined. The Programme, called also 'Safeguarding the National Cultural Heritage', aims to initiate, support and co-ordinate projects for preservation of cultural heritage worldwide. The basic idea is that a cultural society is obliged to collect documents about its own history, preserve them and keep them secure for future generations.

Documents of a great historical and cultural value are often requested for studying purposes. Removing them from optimised climatic conditions in the repository can, however, cause them irreparable damage. Even slight, hardly noticeable damage can result in the loss of such a document. There are numerous examples when even a careful, qualified restoration process turned into a disaster. The primary task is therefore to prevent usage of the original documents. The secondary task is, on the other hand, making the documents accessible by providing acceptable replacements using a state-of-the-art technology.

2. History of the Czech Project

The Czech National Library, along with the private company Albertina icome, were among the very first participants in the Programme and a contract with UNESCO was signed about the development and publishing of a pilot Memory of the World CD-ROM at the beginning of 1993. This CD-ROM contained bibliographic descriptions of 150 of the most valuable manuscripts and old printed books from the extensive collections of the National Library. Along with bibliographic records, over 100 digitised photographs of these documents were included. The entire task should have been completed in only four months from the final decision to a published CD-ROM discs! This was an extreme task but the joint efforts of the National Library and its private partner proved successful, and the product was presented at the UNESCO meeting in Warsaw in May 1993. This was the very first outcome of the UNESCO Programme. It helped to define further tasks and started a discussion on technological aspects of digitisation.

The following two years provided time necessary for discussions and gaining expertise. Finally, the first

BEST COPY AVAILABLE

manuscripts to be digitised cover-to-cover appeared in April 1995. The *Antiphonarium Sedlecense* (Antiphonary of Sedlec), a unique illuminated 13th Century songbook, was selected and digitised using a high resolution digital camera. In addition to ordinary bibliographic information, detailed descriptions were created by a team of experts from the point of view of a historian, musical historian and a codicologist. At the same time, another digitised manuscript — the *Chronicon Concilii Constantiniensis* (Chronicle of the Council in Constance) — was prepared thanks to support from the Foundation of the National Library and its private sponsors. This was the pictorial addendum to the official chronicle of the Council and it was created by Ulrich Riechental in the first half of the 15th Century. This document is quite unusual in showing not only the church officials and the council alone but also the life of ordinary people at that time. Despite its mainly pictorial contents, the full text of the manuscript in Latin was included on the CD-ROM along with its translation into modern Czech. These two products were the first of the *Memoriae Mundi: Series Bohemica*, a series of digital facsimiles of Czech cultural heritage.

As well as the project in the National Library, projects with other institutions were also realised. Collaboration with the State Central Archive in Prague resulted in 1994 in the electronic edition of a full collection of correspondence from the pen of Clemens Lothar Metternich called *Chancellor Metternich and His Times*. Metternich is known as a primary Austrian politician and he played a major role in international development in the first half of the 19th Century. He was in touch with many important personalities of his era such as Napoleon Bonaparte or Charles Ligne. The CD-ROM contains nearly 2000 digitised pages plus an extensive database of other material. Another project with the Museum of Czech Music ended in the full digitising in early 1996 of *The Two Widows*, an opera by a Czech composer, Bedrich Smetana.

3. Preservation issues

There were plenty of questions to be solved during the past years. One of the very basic concerns was to minimise the need to make the original documents accessible to experts for studying purposes. However, the question was 'will experts really be able to live with the electronic facsimiles instead of the original manuscript'? Numerous surveys and tests showed that (after a necessary 'acceptance' period and a minimum training) a minimum 95% of the experts were able to fulfil their research needs by using the electronic document. In fact, all of those interested in the pure contents of the manuscript were satisfied — moreover, usage of the electronic document even lead to discoveries of certain aspects that were never seen in the original work! In addition to that, the experts appreciate the removal of the stress factor experienced previously when they were only allowed to work with the manuscript for a strictly limited time. The only ones unsatisfied were those studying the material aspects like the writing material, inks and colours used etc. — but they will never be satisfied by any replacement if not 100% identical to the original. Anyway, it was possible to lessen the unwanted 'load' on the originals caused by the need for making them available for top researchers to as little as 5%.

4. Access issues

Although preservation is the primary concern in the project, the access issues appear to be even more exciting. When speaking about experts ever allowed to touch a manuscript we are talking about maximum of a dozen of the topmost specialists annually, sometimes even much less. However, the 'electronic manuscript' can be made available not only to specialists but also to students of history and virtually to anyone interested without any danger to the original. This will again completely change the manner of education in this field as the students can be allowed the same opportunities to access the historical documents as their patrons — even though some of the latter will probably not be very happy about this idea.

However, other questions have arisen as more of the Memory of the World projects started to appear worldwide. Probably the most important one was precisely because of the number of projects — as each of the projects used slightly different approaches, they used different media and various software. As a result of this, there are several products in completely different data formats working on different computer platforms. However, preservation should also mean that the new electronic format should even survive the original document. In this case it would not make sense even if the CD-ROM would last hundreds of years, when in just a dozen years there may not be a computer system available to work with the data, not to mention the software part of today's products. A solution to that can only be in adopting a published and well-known data format common for all of the future projects in every library around the globe.

As far as the image format goes, the above concerns lead to a UNESCO recommendation to use JPEG compressed image files for the Memory of the World projects, because of the excellent quality to file size ratio of the JPEG format. Also, JPEG files can be displayed at a reasonable speed and are compatible across a wide range of computer platforms. Again, our surveys showed that it is not possible to define a 'minimum' or 'recommended' image resolution. This is because the necessary resolution is strongly dependent on the document itself (its quality, size, conditions etc.) and can vary over a large range (Ref 1).

However, the digital facsimiles (or digitised books) do not consist only of images. They should also contain descriptive textual information, from a simple pagination to the full text of the document. This meant that we cannot take advantage of common relational databases because they cannot handle full text. On the other hand there are already standards in the library sector defining the contents of such textual data (e.g. AACR2). Following extensive research in this field, the team of the National Library and Albertina icome Praha suggested a common

defined format based on HTML (hypertext markup language) (Figure 1). Within HTML the format defines certain tags that can be used to create a full-text database structure that can be used while searching the database. This format was proposed to the Subcommittee for Technology of the Memory of the World Programme at its meeting in Prague in March 1996 and is currently under evaluation. It is highly probable that this proposal can be accepted as a UNESCO recommendation for the Memory of the World Programme in summer 1996 (Refs 2, 3).

The acceptance of the descriptive format proposal would mean another qualitative step in the Memory of the World. This would imply that all the data created within the Memory of the World will be fully interchangeable and both media and platform independent. Such data can be accessed at the minimum level by any HTML (i.e. World Wide Web) compatible browser like Netscape, Microsoft Internet Explorer, Mosaic etc. However, it is predictable that more specialised software packages will be developed to provide researchers with much more functionality needed for searching in the documents at high speed, fast retrieval of images, image evaluation and so on. As an example, Albertina icome Praha developed *ManuFret*, a specialised manuscript-oriented software package based on our proprietary full-text technology called Fret (Figure 2). ManuFret can input HTML coded data automatically, index them and provide all the necessary features in a very simple to understand environment. ManuFret is available free to all users of digital facsimiles created by Albertina icome Praha within the Memoriae Mundi Series Bohemica project, as well as being part of a knowledge transfer programme to other libraries worldwide.

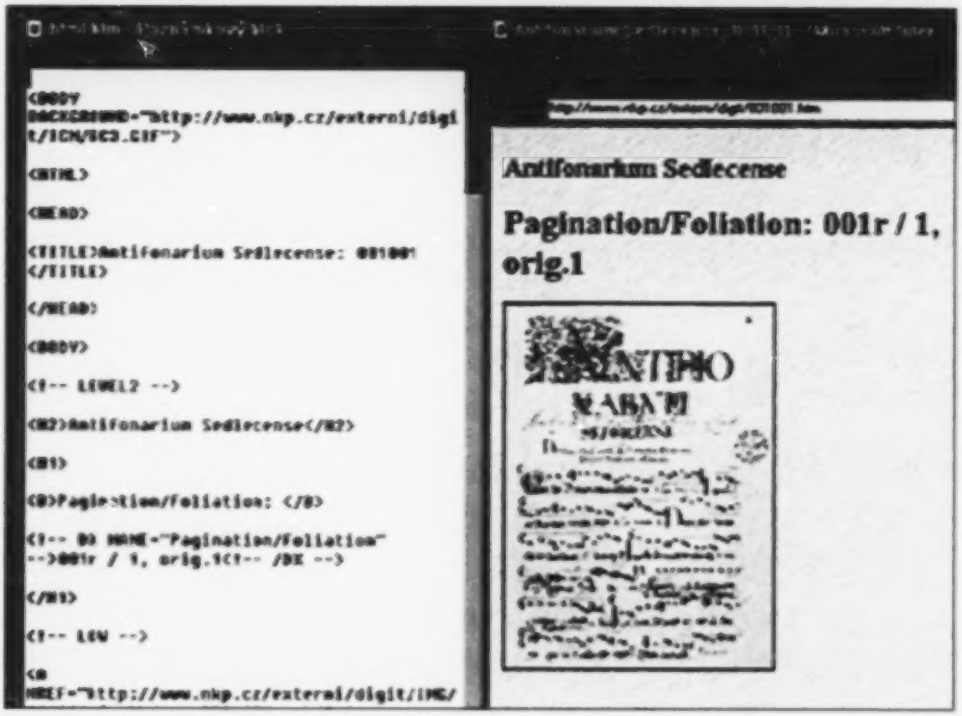

Figure 1: Example of a manuscript's page (Antiphonarium Sedlecense) description in the proposed format: HTML code (left) and its representation in a WWW browser.

BEST COPY AVAILABLE

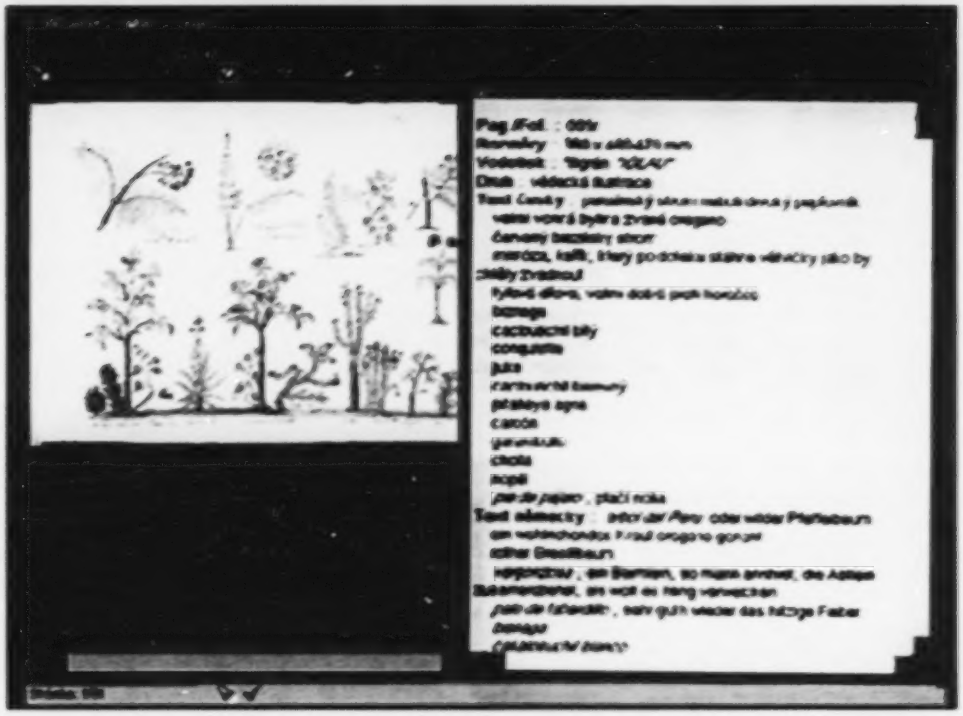

Figure 2: ManuFret software allows for more convenient viewing of the digitised manuscripts. ManuFret provides manual browsing, full-text and fielded searching, adding user notes, marking of selected pages and image handling features. The manuscript shown here is Codex Pictoricus Mexicanus by Ignac Tirsch.

5. Current status

All the projects mentioned lead to a decision to create a large-scale digitisation system in the National Library financed from the national budget, which was approved by the Parliament by the end of 1995. Albertina icome Praha s.r.o. was chosen as a supplier of the necessary equipment and to take over the responsibility for the technical operation of the system. The National Library's responsibilities include selection and handling of the invaluable documents as well as creation of the descriptive contents. The new Kodak 640 high resolution digital camera was chosen for its extraordinary features. Of course, neither flatbed nor drum scanners can ever be used for this purpose as the originals must be handled very carefully and would not withstand excessive opening of the folios or even bending the pages. The system includes two computers (one Apple Mac and one PC with Windows NT operating system) for image input and archiving. The images are finally written to a CD-Recordable disc.

First, manuscripts were scanned by the system just before the submission deadline for this paper. The realistic annual capacity of the system is expected to exceed 15,000 pages. One workshift is reserved for digitisation of documents from the National Library: the outstanding capacity can be used for documents from other institutions.

All the digitised manuscripts are so far available in the manuscript study room of the National Library.

6. Future steps

First of all, we would like to encourage other institutions in the Czech Republic to join the project. There are the Library of the Academy of Sciences, the State Research Library in Olomouc, the Czech Crown Archive and many, many other organisations preserving unique historical documents of great cultural value. Of course we hope to find support for this in both governmental and private sectors, as we did in the past. Although there may be priorities as to which documents are to be digitised first, no one has the right to select which part of the heritage is to be preserved and which not — it must be our task to preserve as much of it as possible.

The catalogue of documents digitised within our project is already being created and is available via the Internet. In the future it will be the task of UNESCO to coordinate the global Memory of the World catalogue. We

BEST COPY AVAILABLE

are also examining the possibility of providing access to digital facsimiles in lower-level quality via the Internet, since the bandwidth is not yet satisfactory for high quality (i.e. large volume) images online. Full quality digital facsimiles on CD-ROM/CD-R are available to other libraries or individuals on demand within the framework of the Memoriae Mundi Series Bohemica Subscription Programme, or on a one-time purchase basis (Ref 2).

Another task will be the establishment of the Memoriae Mundi Series Bohemica Digital Archive. Its aim is to maintain the collection of digital facsimiles created within the project. The Digital Archive will be responsible for storage of all the digital data and, eventually, for their conversion into any future digital format. A high-capacity storage solution will be needed in the more distant future once the decision to go online with the collection has been made.

There are many opportunities for foreign collaboration. Our project is open to any reasonable collaboration and we are ready to help other libraries in starting similar projects with our experience, know-how and software. For example, one such opportunity can hopefully start in collaboration with the Austrian Ministry of Education and the University Library in Graz, Austria.

7. Conclusions

Digitisation of old manuscripts and printed books is coming of age in the Czech National Library. We can foresee years and years of work in order to fulfil the goals of our project. On the other hand it is very exciting and enjoyable to see the project growing from the very beginning, and we are very proud to be at the cutting edge of developing a small but important part of the Global Village.

What would it mean if your library decides to start digitising its entire manuscripts collections?

You would need to:

(1) decide whether your collection is large enough to justify the costs involved in building your own system, or you should better join forces with others;

(2) contact UNESCO GIP to be listed in the Memory of the World Programme;

(3) find the funding for (maybe) a couple of years of work;

(4) learn about existing or *de facto* standards;

(5) test the equipment in order to achieve proper set-up and personal experience;

(6) and then go!

We will be more than happy to help you with any of the above issues.

Vladimír Karen
Albertina icome Praha s.r.o.
Revoluční 13
110 00 Prague 1
Czech Republic
Tel. +42 2 2480 3303
Fax: +42 2 2480 3296
E-mail: aip@login.cz

Stanislav Psohlavec,
Albertina icome Praha s.r.o.
Na drazkách 328
266 01 Beroun 2
Czech Republic
Tel./fax: +42 311 621053
E-mail: aipdev@login.cz

References

[1] Knoll, A. and S. Psohlavec (1995) *Digitization of Old Manuscripts*, Czech National Library, Prague.

[2] Albertina icome Praha WWW home page — http://www.aip.cz

[3] Czech National Library WWW home page — http://www.nkp.cz